

**Computer science education of underserved populations and the efficacy of one-day STEM
events**

Julian Golombek

Georgia Institute of Technology

LMC 4702: Undergraduate Research Thesis Writing

Dr. Amanda Madden

Spring 2021

Contents

Abstract	3
Introduction	4
Literature Review	6
Methods	8
Participants	8
Design	8
Procedure	10
Data Collection	11
Data Analysis	11
Expected Results	12
Discussion	12
Future Directions	13
Conclusion	14
References	15

Abstract

In the U.S., students seeking Computer Science (CS) as an undergraduate major increased by 12% from 2006 to 2012 and then another 51% from 2011 to 2017. Despite this wide pursuit, there is a belief that not enough citizens are being trained to fulfill these positions. Given the opportunity to insert more Americans into this workforce, it should be considered that white males have been the major demographic that have held these secure and high-paying jobs. Georgia's public school system is ranked 34th in the country. In Atlanta, 84% of students are racial/ethnic minorities and 77% are economically disadvantaged. Of the public schools in Atlanta, the average high school graduation rate is 71.1% and the statewide average is 79.4%. Similarly, the average college enrollment rate is 59% and the state average is 64%. This means that only one third of APS students will attend college. HackGT, a CS education and STEM outreach organization, hosts a learn-a-thon for local Atlanta high school students to participate in one-on-one CS learning and mentorship with Georgia Tech CS majors, local professionals, and large technology companies. There has been no official data collection or analyses of the event and its impact has not been validated. Once the restrictions of hosting large in-person events due to the COVID-19 pandemic are no longer in place, the event will resume and two factors will be evaluated. First, further research into how the socioeconomic status and demographic as well as prior perceptions of CS impact future success and involvement in the field. Second, the effectiveness of Catalyst in encouraging underserved high school students to pursue a higher education and career in CS will be gauged and quantified.

Introduction

With the recent growth and incredible dependence on the global technology industry, the demand for skilled workers has been increasing. The technology industry has grown to support almost every other industry, and computer scientists enjoy the benefits of a large job market and higher than average salaries [1]. In the United States alone, students seeking Computer Science and Information Sciences as an undergraduate major increased by 12% from 2006/2007 to 2011/2012 and then another 51% from 2011/2012 to 2016/2017 [2]. Despite this wide pursuit, there is a belief that not enough citizens are being trained to fulfill these positions [2]. Given the opportunity to insert more Americans into this workforce, it should be considered that white males have been the major demographic that have held these secure and high-paying jobs historically. With this in mind, underserved populations from poorly performing school districts can fill the needed computer science positions given the right conditions.

Georgia's public school system is ranked 34th in the country. Of the schools in Atlanta Public Schools (APS), the average graduation rate among APS high schools is 71.1% compared to the statewide average of 79.4%. Similarly, the average college enrollment rate is 59% compared to the state average of 64%. This means that only one third of APS students (33.1%) will attend college. 84% of APS students are racial/ethnic minorities, 77% are economically disadvantaged, and 82% receive reduced price lunches [3-8]. This is a situationally and geographically unique situation, since the Georgia Institute of Technology is also a public university located in Atlanta and many of its undergraduate and graduate programs are nationally ranked in the top ten. This proximity is leveraged by HackGT, a 501(c)(3) non-profit organization that focuses on computer science education and general STEM outreach. The organization, which is operated by Georgia Tech undergraduate students, hosts an annual learn-a-thon for local Atlanta high school students.

Catalyst is a one day event that invites more than 250 students from underserved public schools to participate in one-on-one computer science learning and mentorship with Georgia Tech computer science majors, local industry professionals, and large technology companies and organizations. Overall, the intention is to expose the students to computer science and start them on a path to obtaining a college education and ultimately pursue a job in the technology industry. Since its inception in 2017, the event has been hosted three times with a total of more than 1000 participating students from the greater Atlanta area [9]. Despite the localized achievements of the event and organization, there has been no official data collection or analyses of the event, and its impact cannot be validated.

Given the established event, two concepts will be evaluated. First, further research into how the socioeconomic status and demographic as well as prior perceptions of computer science impact future success and involvement in the field. Second, the effectiveness of the Catalyst event in encouraging underserved high school students to pursue a higher education and career path in computer science and the technology industry will be gauged and quantified.

Test scores are widely used as the primary indicator of students' success. For the past 10 years, low STEM test scores among predominantly minority populations and underserved school zones have prompted investigation and action. New curriculums involving CS classes have been established in schools, but as of 2003, there were less than 50 AP Computer Science teachers for over 370 public high schools [10]. Computer science education has been tackled specifically through programs that target cohorts of demographics to better understand their lack of presence in the field overall. This study aims to investigate a wider breadth of students through the Catalyst event. The socioeconomic status of the students will be evaluated in depth, leveraging short questionnaires with their parents/guardians and a full day of cerebral interaction with

mentors who are trained to take note of their actions. The organization will continue hosting the Catalyst event annually. The study will be expanded each year that the event is hosted. As part of the study, students that attend the events as participants in the study will be followed to track their college and career paths to understand the effectiveness of the event, validate a one-day event, and further validate the studies that the pre- and post-event questionnaires are based on.

Literature Review

In a society increasingly dependent on the services built by people who majored in computer science (CS), the number of graduates in the field and the number of people working in industry both increase annually at an average 20% and 5% respectively in the US alone [11]. Two pertinent issues can be evaluated along the bridge from undergraduate computer science majors to full time employees software engineers - the severe lack of representation of minority and underserved populations and the decreasing happiness of employees. The former can be traced back to initial exposure to the related concepts. The latter is attributed to self-efficacy during undergraduate education. Considering the two proposed problems, the majority of previous studies have examined new ways to expose college-age students to computer science in ways that will actively benefit them once they have entered the field. Research studying the issues independently has not been widely conducted. This includes topics such as introducing high school-aged students to the wider benefits of a specific path and the intent to persist of students that engage in mentorship versus those who do not. The closely related topics can be paired and looked at in parallel with the intention of benefiting both subject groups. HackGT's Catalyst event provides the perfect environment for this analysis and can be evaluated to see if the event's current curriculum achieves its potential benefits.

The benefits of mentorship have been well documented [1,2]. Dr. Ruzica Piskac echoes the prior work done in the field by stating “students seeing a clear impact of their work can have a positive impact on retention.” [2] As described by Dr. Francesca Pernice, mentorship can be used as a tool to predict the trajectories and gauge qualitative characteristics of students [1]. Test scores are the most traditional way to measure a student’s success. It is clearly defined and allows for a straightforward quantitative comparison across the student population. Based on the test itself, a further analysis can be conducted on the number of specific types of questions a student answered correctly or incorrectly. The numerical nature of a test score lends itself to well running statistical analyses and drawing conclusions without the need for further explanation. Despite this, wide scale standardized testing is not a predictor of success indicators such as intent to persist [1]. Defined as GRIT (Guts, Resilience, Integrity, and Tenacity), successfully learning the material taught in school is not the only thing needed to have a successful career. Acknowledging the non-traditional components of education as key points of interest requires the field to develop a way to measure and track them [3,4].

A method to evaluate these indicators that is as robust and definitive as test scores is needed to legitimize such a drastic shift in something as foundational as education. Recently, a CS attitudes instrument was validated by running a voluntary pre-post 5-Point Likert scale based survey for three consecutive semesters in three universities’ introductory computing classes [5]. The validation for the 40 question survey was done by conducting a factor analysis. The question set was from three different studies that were previously conducted. 65% of the questions were from Dorn and Tew’s Computing Attitudes Survey (CAS), which was adapted from Dr. Wieman’s Colorado Learning Attitudes about Science Survey (CLASS) [6,7]. Though validated, 5-Point Likert scales are almost always used to quantify qualitative areas of study. The iterative

turnover of surveys about education forces researchers to validate their methods by leveraging the results of the paper upon which they base their survey. No new instruments are being developed in part because of the difficulty of validating something so qualitative. In line with the area of research, the Catalyst event will be analyzed in a similar manner with the opportunity to develop and verify a new system of evaluation after the study has run for multiple years, giving the participants time to graduate from high school and a significant amount of data to be collected.

Methods

Participants

Since participants will be recruited from Catalyst event, the exact number is not currently known and will not be known until the day of the event and study. Historically, the HackGT organization markets to and plans to have approximately 500 participants in attendance. Thus, it is expected that there will be approximately 300-500 participants in the study. The event will be marketed to high school students in-person and through social media and email. Students will register for the event online (<https://catalyst.hack.gt>). All registrants are automatically accepted. Participants will be able to opt out of the research, as it is not required for participation in the event. Any participant in the Catalyst event is eligible for the study, as the only exclusion criteria is that they must attend the entire Catalyst event to be eligible for the study. If a participant is included in the study and leaves the event before it concludes, they will be removed from the study. A secondary group of participants is the legal guardians of the event participants.

Design

The participating students are from underserved public high schools in the Metro Atlanta and surrounding areas. The study uses multiple questionnaires to collect qualitative information on the immediate impact of the event and perceptions of STEM education from the event participants and the demographic information and socioeconomic status of the event participant's legal guardians.

The event participants will fill out three questionnaires throughout the event. The questionnaires at the beginning and middle of the event will be the same for all event participants. The questionnaire at the end of the event will include two sets of questions, one that is the same for all event participants and one that is specific to the track that they participate in at the event. These two questionnaires will consist of questions with answers on a 5-point Likert Scale ranging from 1-Very unlikely to 5-Very likely and 1-Strongly disagree to 5-Strongly agree as appropriate. The second questionnaire that will be distributed in the middle of the event will be a single question asking the event participants to draw their perception of a computer scientist. Examples of the questions and possible answers are as follows:

(1) Select all of the items that you think you can use computer science to build.

→ A business; Smartphone application; A piece of art; A social media profile; A robot; A community of like-minded people; A website; A computer

(2) For each of the following questions, answer each one to the best of your ability.

→ Using the scale: 1-Very unlikely, 2-Somewhat unlikely, 3-Not sure, 4-Somewhat likely, 5-Very likely

→ I plan on taking computer science classes in the future.; I will spend time learning computer science outside of school.; If my high school offered computer science classes, I would take them.

The questionnaire directly evaluates the immediate success of the event. It measures the participating students' interest and outlook on their computer science. The questions cover their understanding of the field as well as their relationship to it, both academically and career-oriented. The post-event survey, which mirrors the re-event survey, allows for comparison. The net change in answers will be evaluated across all of the event participants and within subgroups such as gender, age, ethnicity, and socioeconomic background.

The secondary group of study participants, the legal guardians of the event participants, will fill out a questionnaire regarding their demographic, socioeconomic status, and education level to be used in conjunction with the event participant data. This study is exploratory, but the comparison of the legal guardians' data with the initial interest in computer science of the students' may shed a light on how social and economic factors may impact one's outlook on academic and professional success.

Procedure

The Catalyst event is a one day event taking place on a Sunday from approximately 8:30 to 21:30, including check-in and check-out. The event itself will run from 10:00 to 20:00. Information and forms (Participant Assent and Parental Consent) will be provided to all participants (event participants and legal guardians) both before the event when they register and at the event. Participants will be able to opt out of the research when they arrive at the event. The first and third questionnaires will be hosted on Qualtrics and accessed on laptops provided to the

participants by HackGT for use during the event. The pre-event questionnaire will be distributed to the participants once they split into tracks at around 10:30. The post-event questionnaire will be distributed to the participants before they leave the event between 19:30 and 21:30. The second questionnaire that will be distributed in the middle of the event will be done on paper. All participants will eat the same meals, watch the same demonstrations, and listen to the same panels. Not all participants will attend the same workshops, as these are determined by the tracks. The secondary group of study participants, the legal guardians of the event participants, will fill out a single questionnaire at the end of the event. All four questionnaires will take between five and ten minutes to complete.

Data Collection

Each legal guardian will receive a parental consent form with two unique matching barcodes. One barcode will be placed on the parental survey. The second barcode will be placed on a removable portion of the participant's event badge. The participants will enter the alphanumeric number of the barcode from their badge into the survey. This will link the two groups respectively to allow for further conclusion to be drawn from the collected data.

Data Analysis

Simple linear regression, standard multiple regression, ANOVA, and MANOVA will be used to determine relationships between demographic information, socioeconomic status, level of education of both the participant and legal guardian, and future prospects of the participant in higher education, specifically in STEM fields. The analysis of this data will continue indefinitely with compounding data as the event is hosted on an annual basis.

In order to establish if the event is immediately beneficial, we are expecting a significant portion of the participants to report an understanding of what computer science is and continued interest or a change from disinterest to interest in studying computer science.

Expected Results

The intention of this study is to evaluate the immediate and long term effectiveness of HackGT's Catalyst event as well as explore the relationship between demographics and perceptions of STEM. It is expected that the event will have beneficial immediate and long term effectiveness. This can be described as establishing and clarifying computer science and STEM education to individuals who may have not known or understood what it entailed, encouraging individuals to pursue a higher education, and displaying a path to achieving a career in computer science.

The Pre- and Post-event surveys mirror each other. Asking the same questions before and after the participants experience the event will help gauge its effectiveness. It is expected that a significant percentage of the Pre- to Post-event surveys will show continued interest or a change from disinterest to an active interest in computer science, suggesting that there is a relationship between the Catalyst event and the possibility of a future in computer science.

Discussion

No data has been collected due to the COVID-19 pandemic which has halted all large-scale in-person events. Once the Catalyst event can take place, it is anticipated that the data collected and results from analysis will align with expectations.

The event has been running for three consecutive years with more than 300 high school students in attendance each year. The event is restricted to Atlanta high school students between

the ages of 14 and 18. HackGT has noted the return of participating students for one to two events as well as the enrollment of participating students at the Georgia Institute of Technology for their undergraduate degrees in Computer Science. It has been additionally noted that some of these students have also joined the HackGT organization to contribute to the planning and hosting of Catalyst events. HackGT markets the Catalyst event to Atlanta Public Schools with low AP class enrollment and high Free and Reduced Lunch enrollment. Since the study is exploratory, it is not known if there will be any significant correlation between parental guardian's demographic, socioeconomic status, and level of education with their child's perception of STEM education, their intention to pursue a higher education, and the idea of computer science as a career path.

We anticipate that the results will verify HackGT's Catalyst event and serve as additional validation for the STAR and Intent to Persist surveys that the questionnaires for this study are based on [1]. Despite the interruption that the COVID-19 pandemic had on the annual hosting of the event, it is expected that the Catalyst event will continue to be hosted for at least eight years. This would allow for eight events worth of data to be collected, improving the compounding analyses that will be conducted. Additionally, it is our intention to initiate an additional component of the study that will track the event participant's education and career paths. With eight consecutive events, enough time will have passed so that participants who attend the event as high school freshmen will have the opportunity to attend college for an undergraduate degree twice over. By the eighth event, those freshmen could be entering the workforce as computer scientists, which is the ultimate goal of HackGT's Catalyst event.

Future Directions

Many extracurricular events exist in the STEM fields, but these are often prolonged exposures, such as after school classes, summer camps, and at home subscription boxes. As Catalyst is a one-day event, the schedule must be targeted and curtailed to be impactful. The efficacy of a single event has not yet been proven, as this study aims to do. It is not feasible for many students, especially those from underserved school systems, to participate in continuous development programs. A single day event is relatively easy to host, as it requires much fewer resources than listed alternatives. If a large standalone event can improve the outlook of the participating students, many more students can be reached. Moving forward, the curriculum and structure of the Catalyst event can be actively changed based on the results of the questionnaires and analysis of the collected data. Ideally starting in the Spring of 2022, the fourth in-person Catalyst will be hosted on the Georgia Institute of Technology's campus. As the first event to officially collect data, the iterative analysis and implementation process can begin.

Conclusion

Pending the lifting of restrictions on gatherings of large people due to the COVID-19 pandemic, HackGT's will resume hosting its annual in-person Catalyst event. Once this comes to fruition, the surveys will be distributed at the events and data will be collected for analysis. Since the event is hosted once per year, the scope of the analysis will increase with the compounding data. It is expected that participating students will become more interested in computer science and pursue a higher education and career in the computer science or science, technology, engineering, and mathematics (STEM) field. The core aspects of the event are expected to remain the same year to year; public high school students from the Metro Atlanta area will be invited to participate in a one-day event that focuses on computer science education. The

specifics of the event, such as the number of students, the activities and curriculum, and the areas in which the participating students attend school may change. Regardless, the data collection and analysis methods will be consistent - a pre- and post-event survey will be used to gauge the participating students' engagement and responsiveness to pursuing an education and career in computer science.

References

1. Baier, Stefanie. 2016. Intent to Persist in College Freshman: The Role of Self-Efficacy and Mentorship. *Journal of College Student Development* 57(5) (Jan 2016), 614-619. <https://doi.org/10.1353/csd.2016.0056>
2. Mark Santolucito and Ruzica Piskac. 2020. Formal Methods and Computing Identity-based Mentorship for Early Stage Researchers. *In Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (SIGCSE '20), (March 2020).
3. Jeremiah Blanchard, Christina Gardner-McCune, & Lisa Anthony. 2019. Dual-Modality Instruction and Learning: A Case Study in CS1. *In Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (SIGCSE '20), (March 2020). <https://doi.org/10.1145/3328778.3366865>
4. Markeya S. Peteranetz and Leen-Kiat Soh. 2020. A multi-level analysis of the relationship between instructional practices and retention in computer science. *In Proceedings of ACM SIGCSE conference* (SIGCSE'20), (March 2020). <https://doi.org/10.1145/3328778.3366812>.

5. Ryan Bockmon, Stephen Cooper, Jonathan Gratch, and Mohsen Dorodchi. 2020. Validating a CS Attitudes Instrument. *In Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (SIGCSE '20), (March 2020).
6. Brian Dorn, Allison Tew. 2014. Empirical validation and application of the computing attitudes survey. *Computer Science Education* 25(1) (2015).
<https://doi.org/10.1080/08993408.2015.1014142>
7. <https://www.physport.org/assessments/assessment.cfm?A=CLASS>
8. Lewis, S., Simon, C., Uzzell, R., Horwitz, A. and Casserly, M. A Call for Change: The social and educational factors contributing to the outcomes of black males in urban schools. The Council of the Great City Schools, Washington, DC, 2010.
9. HackGT Catalyst 2017, 2018, and 2019 Compiled Data.
10. Duda, C. (2014, May 31). A Look Inside Atlanta Public Schools. Retrieved from <https://jjie.org/2012/03/20/by-numbers-look-inside-atlanta-public-schools/>
11. United States, Equal Employment Opportunity Commission. (n.d.). DIVERSITY IN HIGH TECH. New York, New York. <https://www.eeoc.gov/special-report/diversity-high-tech>